EE524 Machine Learning Lab

Assignment 8

18 October 2022

1. Download the IRIS dataset and apply K-Nearest Neighbours (KNN) Algorithm for classifying the samples. Out of the 150 samples, split the data randomly into 80% development data and 20% test data. You can use feature normalization or carry on without it. Take different values of K and train the KNN Classifier on development data, and test it on the test data.
   Implement the KNN Classifier from scratch. Use the Euclidean distance metric and majority voting scheme to decide on the class.

2. For the first question, take K=1 and train the KNN classifier on development data and test it on the same set. What is the accuracy in this case?

3. For the first question, take the values of K from 1 to 30 and see for which value of K, the classifier classifies the samples well in both training and test datasets by considering accuracy as the metric. Plot the K values on the x-axis and accuracy values on the y-axis to choose the optimum K.